Supplemental information

Establishing chromosomal design-build-test-learn

through a synthetic chromosome

and its combinatorial reconfiguration

Jee Loon Foo, Shohei Kitano, Adelia Vicanatalita Susanto, Zhu Jin, Yicong Lin, Zhouqing Luo, Linsen Huang, Zhenzhen Liang, Leslie A. Mitchell, Kun Yang, Adison Wong, Yizhi Cai, Jitong Cai, Giovanni Stracquadanio, Joel S. Bader, Jef D. Boeke, Junbiao Dai, and Matthew Wook Chang
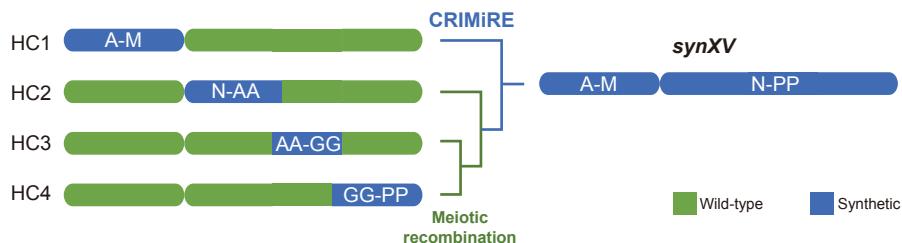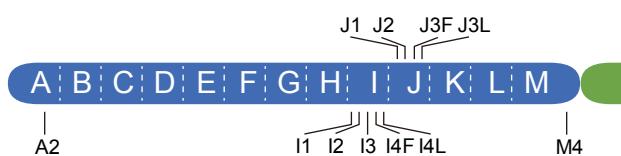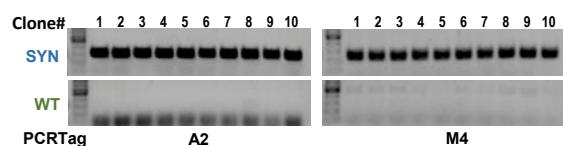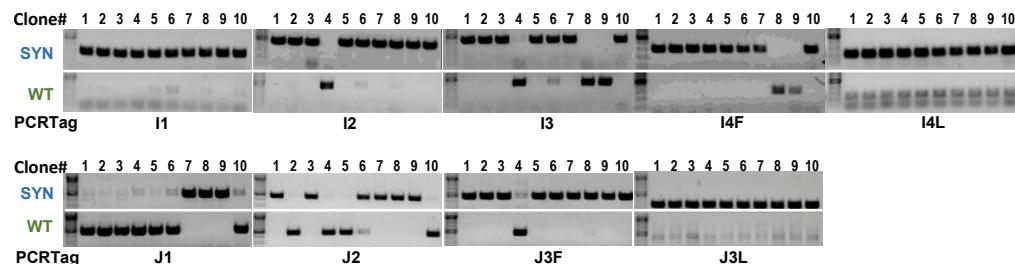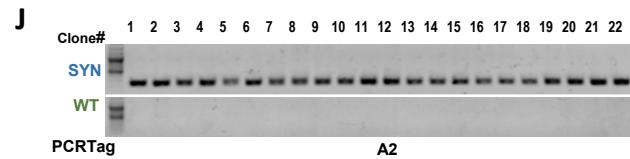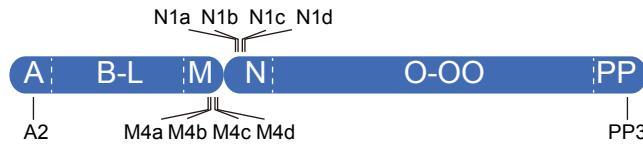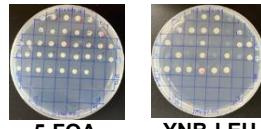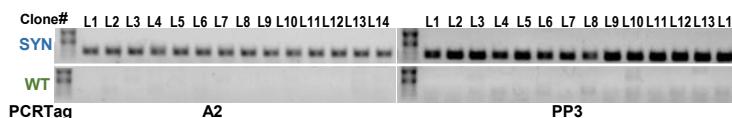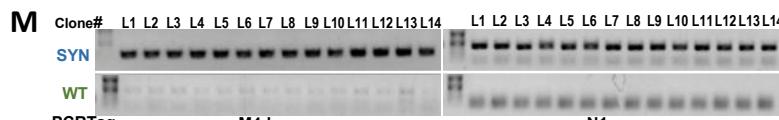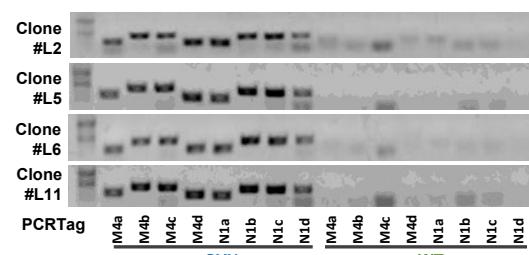
Figure S1.

Figure S1. Assembly of *synXV*, related to Figure 2. (A) *synXV* was divided into four 'hyperchunks', which were constructed concurrently. The semi-synthetic strains (HC1, HC2, HC3, and HC4) were consolidated by meiotic recombination or CRIMiRE to generate *synXV*. Each blue block denotes a synthetic region and the megachunks present are indicated. **(B-E)** Assembly of hyperchunks. Megachunks were sequentially added to replace the native chromosome by SwAP-In. The allocation of megachunks to hyperchunks HC1, HC2 and HC3 are shown in (B), (C), and (D), respectively. HC1 was constructed via consolidation of the semi-synthetic strains HC1A and HC1B by CRIMiRE. (E) HC4 was assembled via several semi-synthetic strains constructed by single-megachunk integration or SwAP-In, and consolidated by CiGA. **(F-H)** Consolidation of HC1A and HC1B into HC1 by CRIMiRE. (F) The approximate positions of the synthetic and wild-type PCRTags (SYN and WT, respectively) used for screening homozygous HC1 clones are indicated. (G) PCRTag analysis of the HC1 candidates at the termini of the hyperchunk using the first PCRTag in chunk A2 and last PCRTag in chunk M4. (H) PCRTag analysis of the HC1 candidates on the left and right sides of the cleavage site. I1, I2, I3 are the first PCRTags of chunks I1-3, and I4F and I4R are the first and last PCRTags of chunk I4, respectively. J1 and J2 are the first PCRTags of chunks J1-2, and J3F and J3R are the first and last PCRTags of chunk J3, respectively. As the distance from the cleavage site increased, so did the probability of having synthetic PCRTags the furthest PCRTags from the cleavage site, i.e. I1 and J3L, were synthetic in all 10 clones. **(I-N)** Consolidation of HC1 and HC234 into *synXV* by CRIMiRE. (I) The approximate positions of the SYN and WT PCRTags used for screening homozygous *synXV* clones are indicated. (J) PCRTag analysis of the *synXV* candidates at the *synXV* termini using the first PCRTag in chunk A2 and last PCRTag in chunk PP3. The heterozygosity was likely due to the proximity of the conditional centromere-*URA3* cassette to the cleavage site, which was <4 kb away. Due to crossover near the cleavage site—which was also observed during HC1 construction—the cassette could have been lost. Consequently, heterozygous strains that could escape the 5-FOA screen for uracil auxotrophy were generated. (K) After performing CRIMiRE with HC1-UL and HC234-U, 14 out of 39 5-FOA-resistant clones were leucine auxotrophs. (L) PCRTag analysis shows that the sequences are synthetic in chunks A2 and PP3 at the *synXV* termini of all the leucine-auxotrophic clones. (M) PCRTag analysis of the *synXV* candidates with the flanking PCRTags nearest the cleavage site. M4d and N1a are the PCRTags in chunks M4 and N1 nearest the cleavage site, respectively. (N) Further PCRTag analysis of chunks M4 and N1 in four *synXV* candidates with PCRTags M4a-d and N1a-d shows the presence of only synthetic sequences.

A
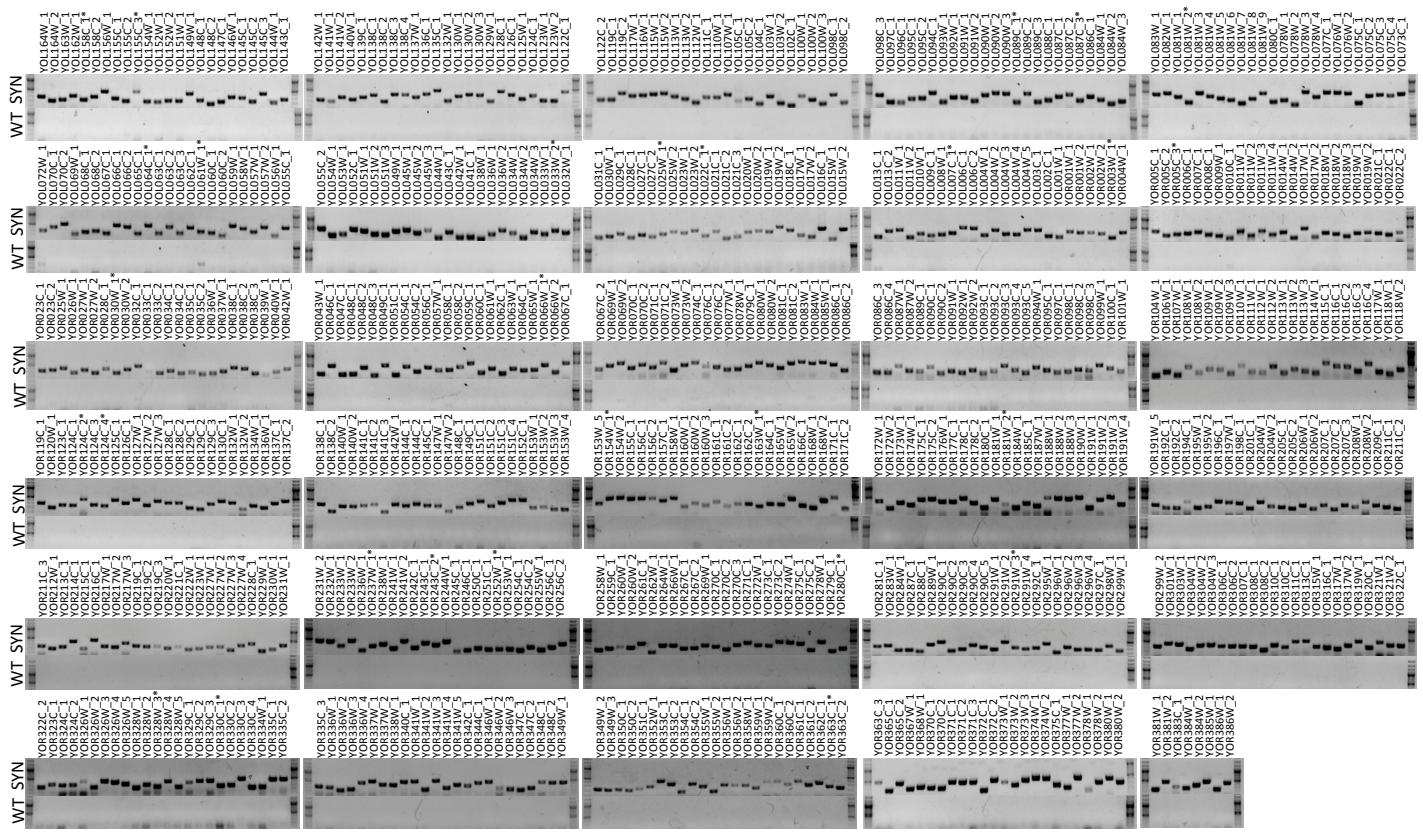
B
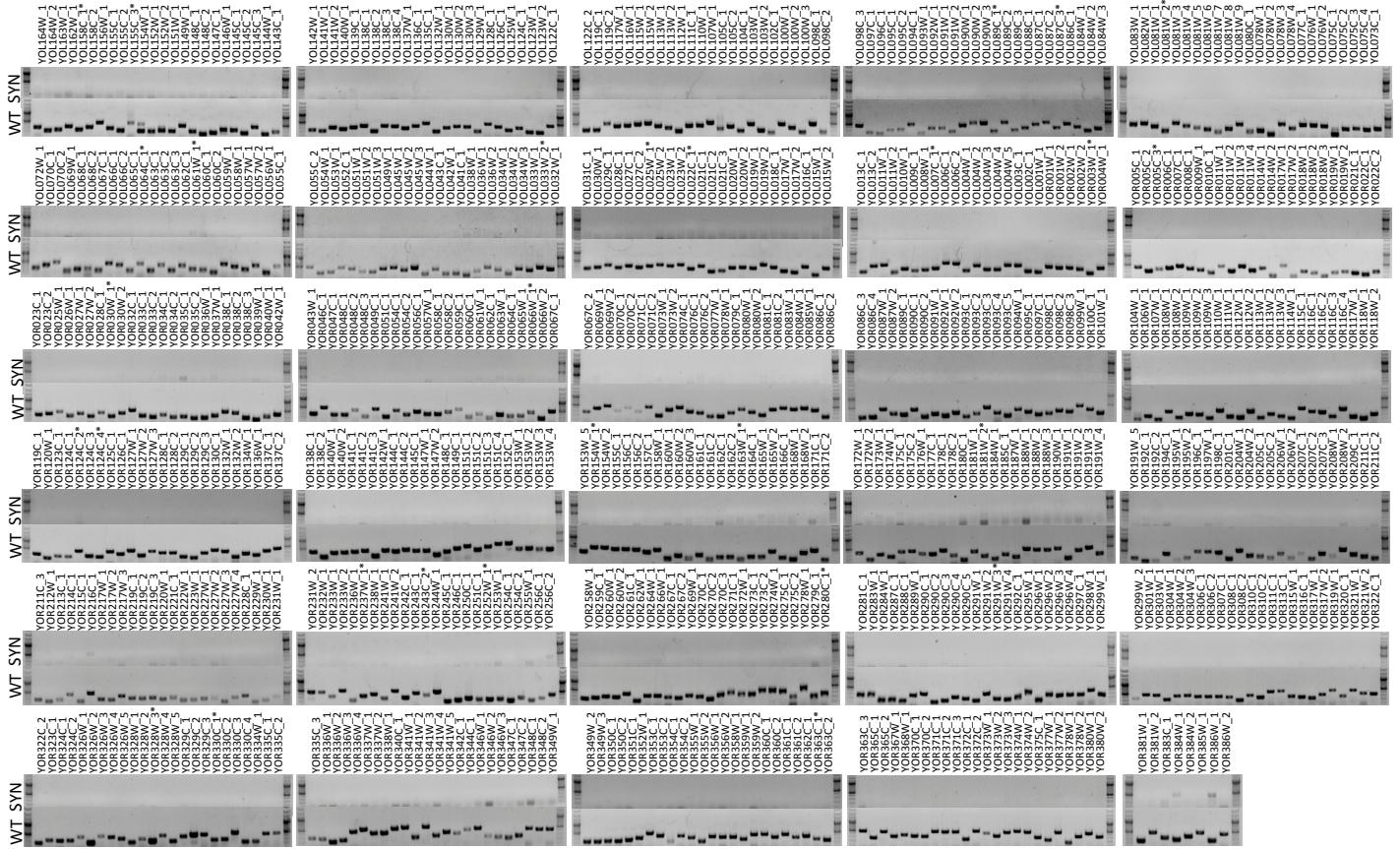
Figure S2.

Figure S2. Complete PCRTag analysis, related to Figure 2. The genomic DNAs of (**A**) synXV_3.1 and (**B**) BY4741 were analyzed using the synthetic and wild-type PCRTags (SYN and WT, respectively). Successful assembly of *synXV* is verified by the presence of all synthetic PCRTag amplicons and the absence of the wild-type PCRTag products in (A). The presence of only the wild-type PCRTag amplicons in (B) indicates the specificity of the wild-type and synthetic PCRTag primers.
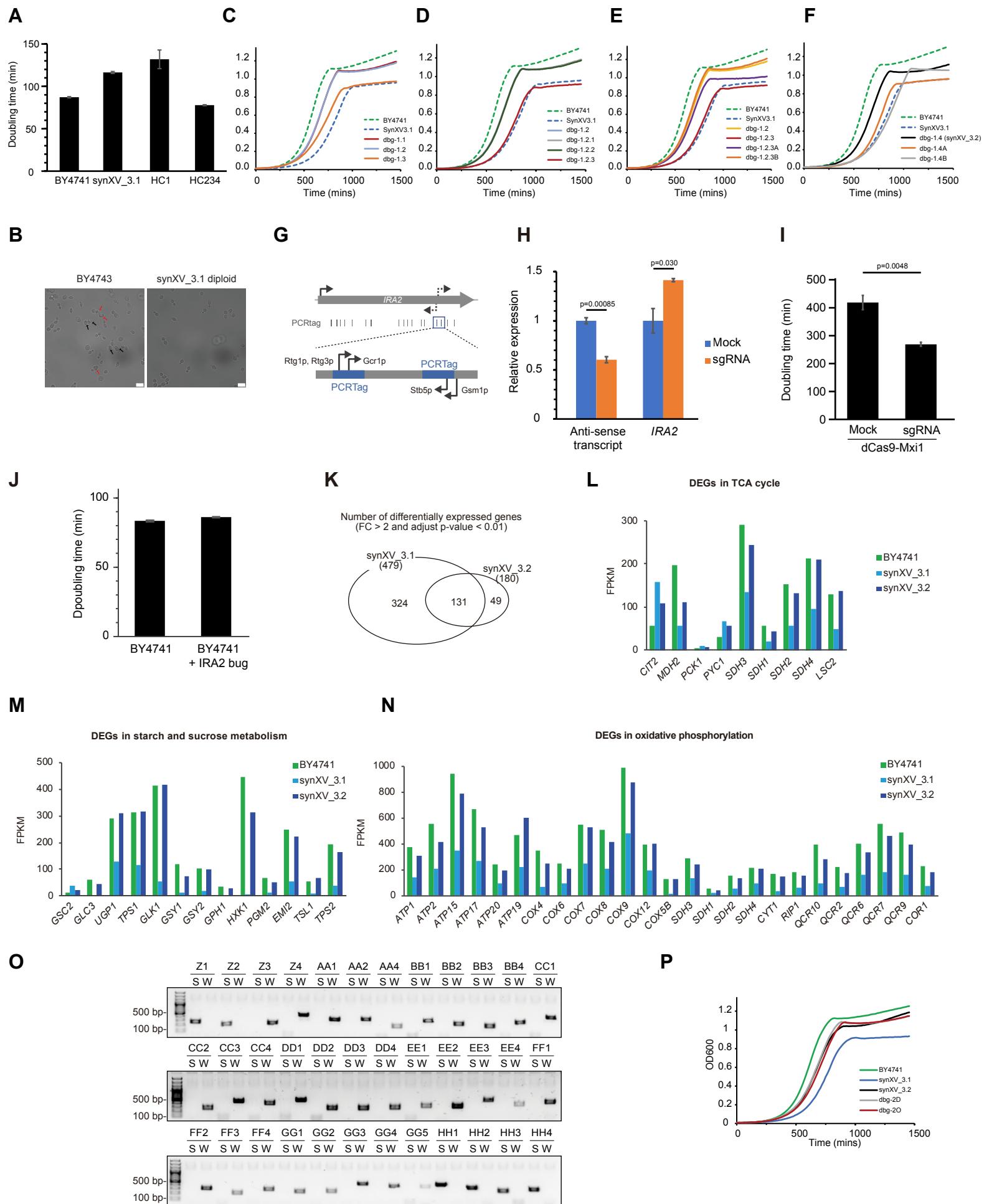
Figure S3.

Figure S3. Debugging of *synXV*, related to Figure 3 and 4. (A) Doubling time of BY4741, synXV_3.1, HC1 and HC234 in YPD (six technical replicates). The error bars represent standard deviation. **(B)** Microscopy images highlight the sporulation defect in synXV_3.1 diploid. Red and black arrows indicate tetrad and dyad cells, respectively. White bars represent 10 µm. **(C-F)** Growth profiles of the strains are shown in Figure 3E. **(G)** Transcription factor binding site prediction by YEASTRACT+. Arrows indicate the direction of transcription by the predicted transcription factors. **(H)** qRT-PCR results showing the repression of the anti-sense transcript of *IRA2* and the concurrent elevation of *IRA2* expression in synXV_3.1 due to CRISPRi. The error bars denote standard errors (calculated from three biological replicates). The p-values were calculated using a t-test. **(I)** Growth assay outcomes showing a reduced doubling time for synXV_3.1 achieved through CRISPRi in YPD medium (200 µg/mL G418 and 200 µg/mL hygromycin B) at 30°C. The error bars denote standard errors (calculated from three biological replicates). The p-values were calculated using a t-test. **(J)** Growth assay outcomes demonstrating that the defective region of *IRA2* does not induce deleterious effects in a wild-type background. The growth assay was conducted in YPD medium at 30°C. The error bars denote standard errors (calculated from three biological replicates). **(K)** Venn diagram showing the number of differentially expressed genes in synXV_3.1 and synXV_3.2 compared to BY4741. FC, fold change. **(L-N)** Bar plots showing that the FPKM of differentially expressed genes (DEGs) in synXV_3.1 were enriched in the TCA cycle (L), starch and sucrose metabolism (M), and oxidative phosphorylation (N). The FPKM values were averaged from biological triplicates. **(O)** The genomic DNA of mr-wtAA-GG diploid was analyzed using a synthetic and a wild-type PCRTag (S and W, respectively) from each chunk for megachunks Z-HH. **(P)** Growth profiles of dbg-2D, dbg-2O compared to BY4741, synXV_3.1, and synXV_3.2.
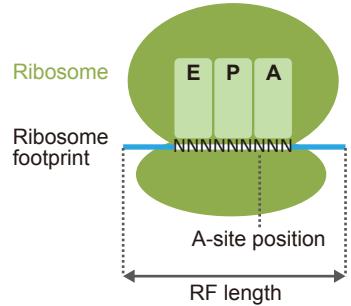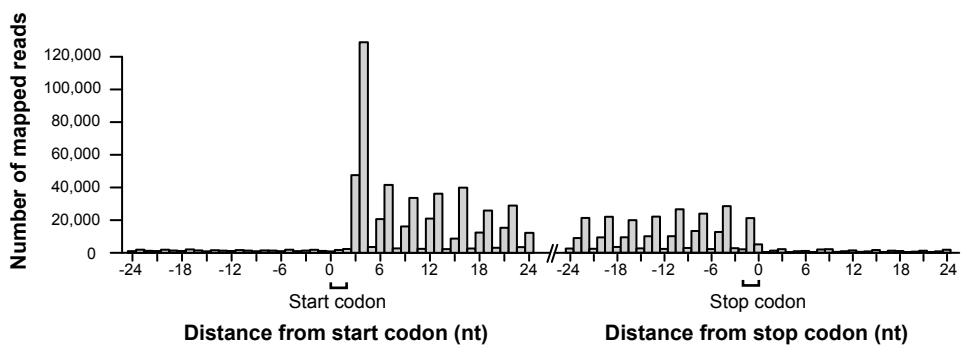
**Figure S4**

Figure S4. Confirmation of accurate A-site peak prediction and comparison of relative ribosome occupancy (RRO) value distributions between BY4741 and synXV_3.1, related to Figure 5. (A). Schematic drawing explaining how the A-site positions were predicted from the ribosome footprints (RFs). **(B)**. Predicted A-site locations aligned to start codons or stop codons showing that the peaks were observed on stop codons but not start codons. **(C)** The bold, bottom, and top lines of each boxplot represent the median, 25th percentile, and 75th percentile of RRO, respectively.

A

ΔG of WT 3' UTR (kcal/mol)
ΔG of SYN 3' UTR (kcal/mol)

B

Readthrough efficiency (RE)
= Readthrough read (RPKM) / CDS read (RPKM)

CDS read   Readthrough read

BY4741

synXV_3.1

99 nt

C

Readthrough efficiency

BY4741

synXV_3.1

Relative gene position on chrXV or synXV

D

REV1   [0 - 2.00]

BY4741 (Ribo-seq)

[0 - 2.00]

synXV_3.1 (Ribo-seq)

YOR345C

E

REX4   [0 - 6.00]

[0 - 6.00]

YOL079W

F

SFL1   [0 - 2.00]

BY4741 (Ribo-seq)

[0 - 2.00]

synXV_3.1 (Ribo-seq)

YOR139C

G

OD₆₀₀
Time (mins)

BY4741
synXV_3.1
synXV_3.4

Figure S5

Figure S5. Effects of LoxPsym insertion on translation around 3' and 5' UTRs, and the growth profile of synXV_3.4, related to Figure 6. (A-C) 3' UTR insertion of LoxPsym did not increase translational readthrough. The 3' UTR of genes on *synXV* was analyzed by computing the free energy of 50 nt after the stop codons using RNAfold. (A) The scatterplot shows the free energy (ΔG) reduction in the 3' UTRs of genes on *synXV* due to the strong secondary structure of the inserted palindromic loxPsym. In theory, the strong secondary structure might cause translational readthrough, a phenomenon known to depend on the sequence and secondary structure of the 3' UTR [1]. Such occurrences of translation past the stop codon could result in the translation of C-terminal elongated proteins with different properties or even dominant deleterious effects [2]. We computed the readthrough efficiency (RE), as defined in (B). (C) Scatterplot showing the REs of *synXV* genes. None of the genes with 3' UTR loxPsym insertions in synXV_3.1 were found to have significantly higher REs than those of the respective genes in BY4741. Hence, we concluded that the 3' UTR loxPsym insertion does not elicit translational readthrough. In line with the lack of 3' UTR-related bugs in previous Sc2.0 studies, the 3' UTR loxPsym feature in Sc2.0 per se is likely innocuous despite its extensiveness. **(D-F)** Translational repression by loxPsym insertion into 5' UTRs and growth profile of synXV_3.4. Ribo-seq data showing the accumulation of ribosome footprints around the loxPsym at the 5' UTRs of *REV1* (D), *REX4* (E) and *SFL1* (F). The black boxes and red boxes represent coding regions and loxPsym sites respectively. White arrows indicate the direction of genes. **(G)** Growth profiles of BY4741, synXV_3.1 and synXV_3.4. in YPD at 30°C.
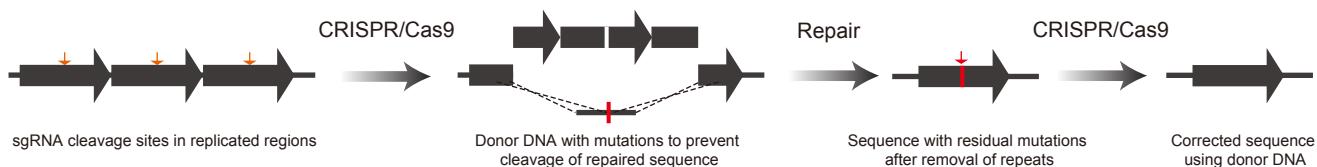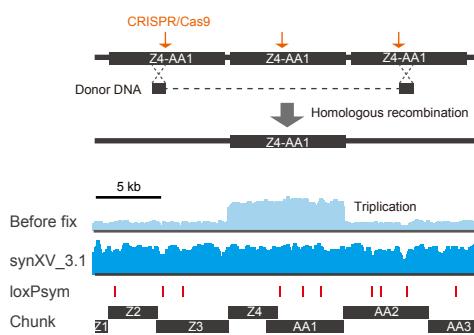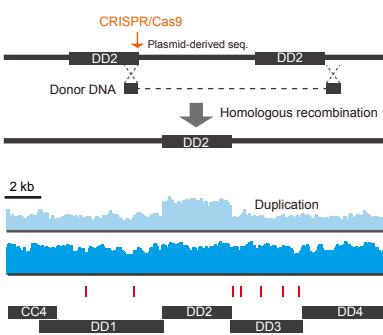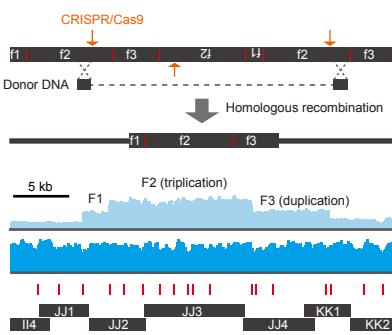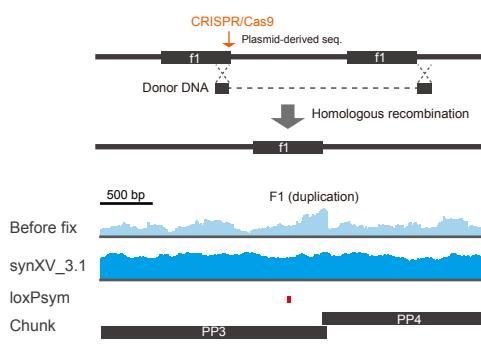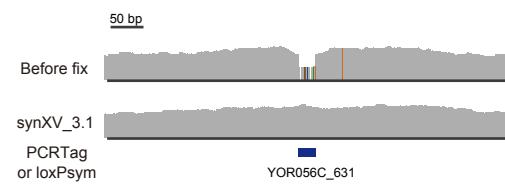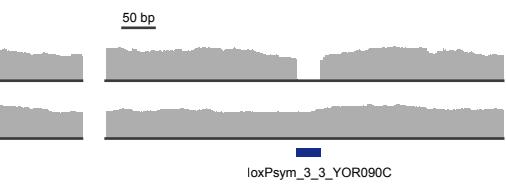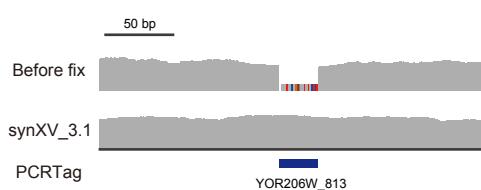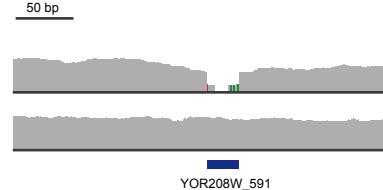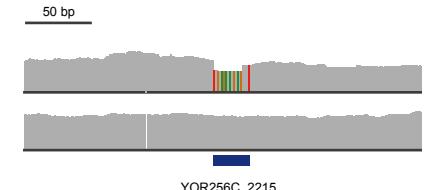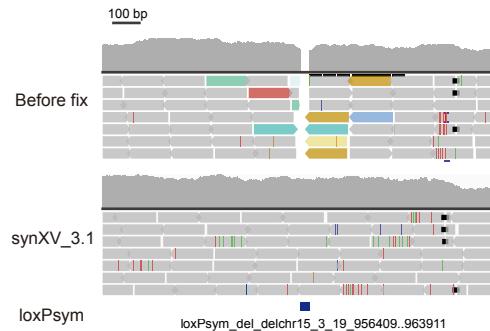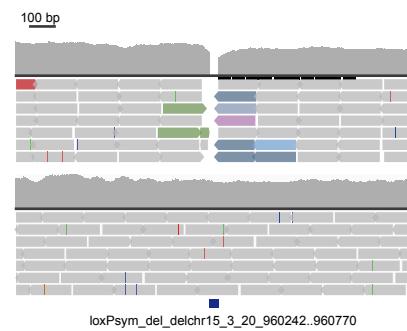
Figure S6

Figure S6. Fixed replicated regions and sequence deviations using CRISPR/Cas9, related to STAR methods. (A) Schematic drawing of two-step replication repair. **(B-E)** Schematic illustration of the methods used to remove replicated regions and whole genome sequencing data showing successful removal of the replications. The arrows indicate the CRISPR/Cas9 cleavage sites. **(F-J)** Whole genome sequencing data showing successful incorporation of PCRTags (F, H-J) and loxPsym (G) into synXV_3.1. **(K, L)** Whole genome sequencing data mapped to *synXV* sequence showing the successful incorporation of loxPsym sites into synXV_3.1 to replace repeat elements within chunk MM1. The existence of Ty elements before fixing at these loci due to wild-type sequences are revealed by the colored bars, which represent paired reads mapped to Ty elements; the full-length repeats are not depicted in the figures because the reads were mapped to the *synXV* sequence.